CHAPTER 5. CONTROL

This chapter deals with two related issues. First, the *comparability* of cases. Second, how it is possible to reduce, and possibly eliminate, the influence of "third variables" on the relationship between independent and dependent variables under investigation, that is, variables the researcher wants to *control* for. Classification and taxonomical treatment underlie both issues. For this reason they have been grouped together in a single chapter.

Comparability: The Limits of Comparison

Comparative researchers are often confronted with the problem of what is and what is not comparable. This relates to (1) the *limits* of comparison and (2) the *strategies* for making cases comparable.

Are there *logical limits* to comparison? Are there things that are "too different" to be compared and to be included in the same research design or, on the contrary, is "everything comparable?" Is it possible to compare the election of a U.S. president with the selection of the head of a tribe in the Amazonian jungle? Questions of comparability have important consequences. A *historicist approach* tends to consider phenomena as *unique*. It is the supposed "uniqueness" of events that makes them noncomparable. The critique to extended comparisons takes sometimes the form of "ethnocentrism," which maintains that concepts developed in the frame of a given society/time do not fit other cases, and that what we know about some societies cannot be extended to others.

As seen in the definition in Chapter 1, however, with the comparative method we do not compare directly two or more cases. *We compare the values that the common attribute takes for each case*. Let us take, as examples from different subfields in the social sciences, the following statements:

- Crime is higher in suburban areas than in inner city centers.
- Welfare policies of the new government are more restrictive than those of the previous government.
- The electoral system of Brazil is more proportional than the Argentinian system.

First, for each statement there are two objects: urban areas, governments, electoral systems. Second, for each statement there is an attribute that the two objects share: crime rates, welfare policies, the proportionality of

the electoral system. What is compared between suburbs and city centers is the rates of crime (operationalized, say, as number of reports to the police). In the second example, levels of "generosity" of welfare policies can be measured through health benefits, pensions, unemployment subsidies, and so on. In the third example, we do not compare electoral systems as such but their proportionality. The Gallagher least square index of disproportionality (Lsq) is 3.70 in Brazil and 13.5 in Argentina. The electoral system in Brazil is therefore more proportional than the Argentinian one.

The question of comparability is therefore one of *sharing common attributes*. What is compared are the values of the cases on these shared properties. From a methodological point of view, therefore, *there are no limits to comparison*. It is possible to compare the duration of "office" between the U.S. president and the head of a tribe. In the former case it is 4 years (renewable once), in the second it is lifelong. We can also compare their selection: by election and birth. When comparing the values of the common attribute (duration of office, method of selection) the level of measurement might be nominal, ordinal, or quantitative. In the comparison of welfare generosity values are quantitative; but in the comparison of the selection of heads of state, values are nominal.

Taxonomical Treatment

The comparability between cases is thus given by *sharing a same attribute or property*. If cases A, B, . . . N share the attribute X, then their values (0, 1, 2, etc.) can be compared. Comparability is obtained by finding a "common denominator" between cases. As Sartori (1970) has put it, "to compare is 'to assimilate,' i.e., to discover deeper or fundamental similarities below the surface of secondary diversities" (p. 1035).

The first step in the logical control procedure is therefore conceptual and consists of defining *empirical universals* making cases comparable (Sartori, 1970, 1984a, 1991). Empirical universals are concepts or categories defining the attribute shared by the cases compared. The transformation of single historical observations into comparable cases is obtained—in a famous phrase—through "*the substitution of variables for proper names*" (Przeworski & Teune, 1970, p. 25; see also Collier, 1991a, 1991b).

This is why a purely "case-oriented" approach is not tenable. Comparison implies variable-thinking, attributes, properties. If this is missing, no comparison is possible: "the rule of holism yields a clear and straightforward contradiction: only incomparables are comparable" (Zeldich, 1971, p. 276). As Bartolini (1993) notes, "[c]ases cannot be compared 'as wholes,' but only when common properties are identified"

(p. 137) (see also Goldthorpe, 1997a, pp. 2–4). Both the so-called "variable-" and "case-oriented" approaches—one relying primarily on statistics and large-N designs; the other on Mill's methods, Boolean algebra, and small-N designs—reason in terms of variables and are interested in variable analysis. This again stresses similarities between the two traditions of comparative studies.

Classification and Typologization

Classification allows one to establish which cases are comparable, that is, share a common attribute (Kalleberg, 1966). By establishing what is similar, what belongs to a same group of cases or class, one establishes whether or not they share an attribute and can be compared.

1. *Equivalence*. Comparable means something that shares a same attribute (electoral turnout, animist rituals), that is, belongs to a same class of cases. If we wish to study electoral turnout in national parliamentary elections, we must first be able to establish which countries can be included and which cannot. If we wish to study animist rituals, we must first be able to establish which cultures have animist rituals and which do not. Comparable is something that has a given degree of "sameness," something that belongs to a same group defined by a shared attribute. To be able to discriminate in that way one must define clearly what is meant by "electoral turnout" or "animist ritual." The concept or category must have the same meaning for all the cases included in the comparison.[13] The category must be *equivalent* (van Deth, 1998). Say, by "electoral turnout" we mean voting in free, recurrent, and correct elections by universal suffrage for a parliament in which more than one party present lists and candidates, and with alternative sources of information. In such a definition one would not include China today. China is therefore not a comparable case.[14]

Before one can investigate the presence or absence of attributes, or before one can rank objects or measure them in terms of some variable, one "must form the concept of that variable" (Lazarsfeld & Barton, 1951, p. 155). The concepts or categories should never be *vague*, that is, they should always indicate *to which empirical aspects they refer*. Empirical *referents* of concepts should always be clearly stated: "by electoral turnout we mean [a number of empirical referents]." Only in this way it is possible to say whether or not a case is indeed a case of turnout. In other terms, it is only through nonvague concepts that one can establish if cases share the same attribute—and ultimately establish their comparability. If the meaning of this concept or category is precise, its *discriminating power* is enhanced, that is, it divides the domain of cases into classes separated by a sharp boundary.

This has important consequences for data collection. Concepts and categories are "data containers" and these should be defined in such a way as to increase their discriminating power, that is, clearly indicate which are cases of turnout and which are not. One can compare levels of political violence in Chile and in Canada only if the same thing is meant in the two cases. One cannot compare political violence in these two cases if in the former case one includes as empirical referents killings, kidnappings, and street violence and in the latter sit-ins, manifestations, and verbal attacks to political leaders.

2. *Logic of classification*. Classification is the most important procedure of *concept formation*: "[it] is the basic type of concept-formation in science. Neither comparison (nonmetrical ordering) nor measurement proper can take place without it . . . Comparison can only be made after classification has been completed" (Kalleberg, 1966, pp. 73, 75). "In short, *two objects being compared must be of the same class*—they must either *have* an attribute or *not*. If they have it, and *only* if they have it, may they be compared as to which has it more and which has it less" (p. 76).

Following Bailey's (1994) definition (in this same series), classification is a general process, as well as the result, of grouping cases in terms of similarity. In establishing groups and classes, we want to minimize differences within each group while maximizing differences between groups. The similarity element defines which objects belong to the same class (genus), whereas the difference element defines what distinguishes classes (species and subspecies).

The classificatory treatment of concepts has three basic rules.

- *Dimensions of classification*. Classifications are based on explicit criteria for the creation of groups. Groups can be based on a single dimension or property (one-dimensional) or based on a number of dimensions (multidimensional). The term *typology* is used for multidimensional classifications in which categories are distinguished conceptually rather than empirically (*taxonomy*).
- *Mutual exclusiveness*. There must be only one class for each item. No case should belong to more than one class. If a classification has a mutually exclusive set of classes, they do not overlap with each other.
- *Joint exhaustiveness*. There must be a class for each case. No case should be left out. If the categories are exhaustive, each case will be

in one of the categories of a variable. A problem sometimes is that, to have a class for each case, classifications tend to have a high number of classes. To avoid that sometimes categories such as "none" or "other" are included.[15]

It is often maintained that one of the specificities of the small-N comparative method is the more extensive use of classification. Its role in comparison is indeed crucial. However, the role of classification is equally important in other methods, namely the large-scale comparison based on quantitative variables and statistical techniques.

Classification precedes statistics; it is not alien to it. Concept formation refers to differences in *kind* rather than *degree* (Sartori, 1970, p. 1036). The taxonomical hierarchy from more general to more specific touches directly on problems of membership in categories and *classification* principles. The *qualitative* logic of *classification* therefore comes before that of *order and quantity*. The logic of gradation belongs to that of classification. Before being able to use the signs "more than" (>) and "less than" (<) one must establish the signs "equal to" (=) and "different from" (≠). Comparability is therefore a problem of "what," a qualitative problem that cannot be replaced by the question of "order" or "how much."

Levels of Generality

1. *Conceptual stretching*. If concepts are able to "travel," then they apply to a large number of comparable cases. However, not all concepts and categories are good at traveling: some are developed in the frame of specific geographical, cultural, and socioeconomic contexts and, when extended to "new" cases, they do not make sense. This problem is particularly acute in cross-national studies. "Western concepts" have a different meaning in other parts of the world. What Sartori (1970) has called the "traveling problem" (p. 1033) is closely related to the "expansion of politics," that is, an objective increase of the number of cases and a subjective increase of interest in sociopolitical issues since the 1960s.

The traveling problem appears when concepts and categories are applied to new cases, different from those around that they had originally been developed. A frequent but inadequate answer to this problem has been "conceptual stretching" (Hempel, 1952; Peters, 1998, pp. 86–93; Sartori, 1970). *Conceptual stretching* refers to the distortion of concepts to make

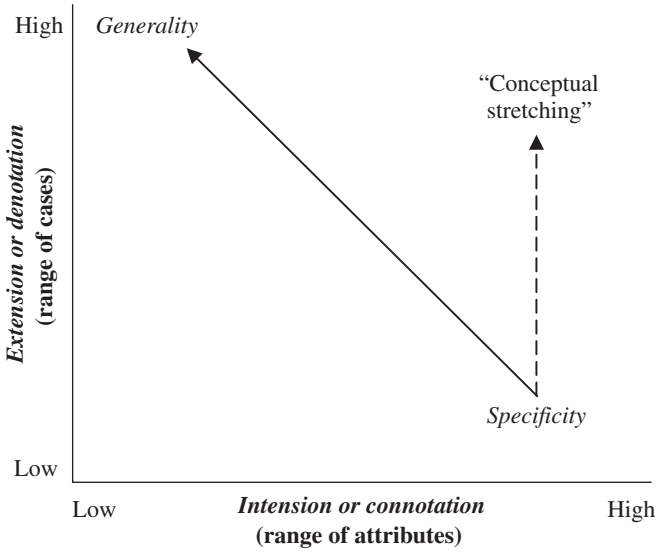
them fit new cases. The stretching arises when *a category developed for one set of cases is extended to additional cases and these new cases are too different and the category is no longer appropriate in its original form* (Collier & Mahon, 1993).

2. *Ladder of abstraction*. How can conceptual stretching be avoided? First, comparative research relies on *empirical universals*, or *observational concepts*, that is, abstractive inferences from empirical observations rather than *theoretical (nonobservational) concepts* such as "system," "feedback," or "equilibrium." These concepts have no empirical referents. They are nonoperationalizable, that is, nonmeasurable.

Second, if we want to increase the number of cases, to avoid conceptual stretching we must simultaneously decrease the characteristics and properties of the empirical concept. This is done by climbing the so-called "ladder of abstraction" (Sartori, 1970, p. 1041, 1984a, p. 24) or "ladder of generality" (Collier & Mahon, 1993). Empirical concepts can be placed at different levels of an imaginary scale. Their vertical positioning on the ladder depends on the relationship between *intension and extension of the concept*.

- *Extension (or denotation):* These terms indicate the set of objects, phenomena, events, or entities to which the concept or category refers. The extension of a concept is the class of "things" to which it applies.
- *Intension (or connotation):* These terms indicate the set of attributes, properties, or characteristics of a concept or category. They define the category and therefore determine the membership of a case to it. The intension of a concept is the class of properties that determine the "things" to which the concept applies.

The relation between extension and intension obeys a *law of inverse variation*: the greater the intension of a concept, the more limited the "things" that belong to this class as defined by the attributes of the concept (Collier & Mahon, 1993). In other words, the richer and longer the list of characteristics of a concept, the more limited the set of objects to which this concept applies. Conversely, the more limited the specification of attributes and properties of a concept, the larger the class of "things" (entities, objects, events) to which the concept refers.



It is possible to compare the selection of the executive in Germany and in the United States insofar as the attribute "selection" applies to both cases. These two cases are comparable insofar as they share an attribute. However, it is not possible to compare the "election" of the executive between these two countries, as in the former the executive is nominated by the parliament and not elected. Comparability therefore depends on the *level of generality* of the language that is applied to observations. "Selection" is more general than "election."

A limited intension of the concept of "executive selection" without specifying attributes covers a larger number of cases. On the contrary, a larger intension, that is, more specifying attributes—such as "direct election by voters"—would exclude a number of cases in which executives are either appointed by the parliament (most European democracies) or are elected indirectly through an electoral college (as in the United States).

There are two ways of climbing the ladder of abstraction. One is by broadening the extension of a concept (a reduction of attributes or properties). The result is a larger class with lesser differentiation but still with clear boundaries and discriminating power. This is the correct way of proceeding. The other is "conceptual stretching," which increases the extension *without* diminishing the intension. The extension is increased by obfuscating the boundaries of the concept (Sartori, 1970, p. 1041).

Family Resemblance and Radial Categories

In discrete categorization, cases either belong to a class or not: they either have an attribute or not (Kalleberg, 1966, p. 76). As Sartori noted, however, the requirement of *positive identification of attributes* may be, in practice, too exacting. His answer to the problem was to say that, when it is not possible to define the exact attributes of a concept, then one must state clearly what the concept *is not*, that is, *a negative identification of attributes* (Sartori, 1970). Two further solutions to the problem posed by discrete boundaries have been proposed: (1) *family resemblance* and (2) *radial categories* (Collier & Mahon, 1993).[16]

1. *Family resemblance*. Family resemblance (originally developed by linguistic philosophers such as Wittgenstein) is based on the principle that, if there is no single attribute that all the members of a category share, researchers can include cases that share the attribute to *varying degrees*. Take, for example, "democracy" defined through

> (1) universal suffrage (political rights); (2) free press, association, belief, individual's protection (civil rights); (3) free, recurrent, correct elections; (4) executive responsible before legislative; and (5) independent judiciary.

If we compare Britain, France, Germany, and Belgium in the late 19th century we see that the attribute "democratic" is not perfectly shared by all cases. In Britain there was no universal suffrage, in France the judiciary was not independent, and in Germany the government could not be outvoted by the parliament. With classical categorization only Belgium would fall into the "democracy" category.

The idea of family resemblance is to consider that the attribute is shared to *some degree* by all the cases. The *prototype category* (democracy) is an *analytical construct* with a heuristic usefulness. Max Weber's ideal types are categories that were defined analytically rather than based on attributes shared by empirically observed cases (Burger, 1976). Real cases share with the ideal type its defining attribute to some degree, meaning that the attribute assumes a "varying geometry" across cases. The advantage is that useful categories are not abandoned hastily by being overly strict.

2. *Radial categories*. Radial categories (originally developed by cognitive scientists such as Lakoff) also rely on a "varying geometry" of attributes across cases.

There is a "prototype" or ideal type representing the perfect or more complete case. This is the "primary subcategory" of which "secondary subcategories" are a variation. Secondary subcategories do not include all

the properties of the primary one. Noncentral subcategories arise when the component elements of the primary subcategory are taken one-by-one or in different combinations (but not complete). It is different from classical categorization where there is a progressive differentiation into genus (superordinate) and species (subordinate). What differentiates the superordinate category from subordinate categories is that the subordinate categories have *more* properties that are added to the superordinate one to differentiate different types (are more "precise"). In radial categorization what differentiates the secondary from the primary subcategory is that we have *less* component elements.

These classification strategies offer different answers to how we construct categories. This has consequences on comparability, that is, which cases we include in the analysis.

Control and Research Design

The previous section has discussed the importance of classification and taxonomic treatment for matters of *comparability*. Classification has a second important role. Classification allows one to *control for variables* (Smelser, 1976, 167–174). These two roles should not be confused. Comparability concerns *cases*; control concerns *variables*. Once comparability has been established, classification becomes an instrument to exclude factors researchers do not want to influence the relationship under investigation—a process through which unwanted sources of variation are reduced.

Matching and Randomization

Empirical research is based on hypotheses concerning causal relationships between phenomena (or variables, once they have been operationalized). Through tests against empirical evidence hypotheses are either verified or rejected. The empirical test of hypotheses implies two separate but related aspects:

1. determining the *association* between phenomena, that is, between cause and effect (in operational terms the *association between independent and dependent variables*);
2. while *isolating* it from the influence of other variables to establish—one by one—the causal role of each operative variable independently.

The same variable can be an experimental or control variable in different phases of the test depending on whether or not it is "allowed" to vary.

Through the transformation of independent variables in control variables and vice versa for all variables that are considered relevant, the investigator progressively gains confidence in the explanation, refines the theoretical argument, and strengthens results. As for all types of research, which variables should be controlled for is a decision by researchers themselves based on previous knowledge (theory) or new insights.

The control of variables can be carried out through

- *Randomization (control through MDSD).* With randomization *differences* are excluded: if the same phenomenon occurs in different contexts, it follows that differences do not account for its presence and, thus, are irrelevant. This is similar to MDSD and, as will be seen, to the Method of Agreement. The MDSD eliminates "third variables" for which values vary across cases.
- *Matching (control through MSSD).* With matching *similarities* are excluded: a variation in the dependent variable cannot be caused by a factor that is constant across cases. Through matching the influence of third variables is excluded by transforming them into constants and thus do not represent an unwanted source of variation. This corresponds to the MSSD and to the Method of Difference. The MSSD eliminates "third variables" for which values are constant across cases.

To *randomize* means to select cases that cover the entire range of values of a given property. Random samples, which assure that each case in the universe (or population) has an equal chance of being drawn, allow one to infer with more confidence. Randomization processes are typical of statistical methods, which can rely on large numbers of cases. To *match* (sometimes called *parameterization, standardization,* or *stratification* in the case of sampling) means to transform variables into *constant* scores that do not vary so that their influence is excluded and the relationship between independent and dependent variables is isolated.

It is important to note that both randomization and matching as techniques for controlling unwanted sources of variation rely on procedures of *case selection*, that is, ultimately on the *research design*.

In the social sciences research designs are particularly important because researchers draw cases from *already existing* data. In experiments, investigators have a direct influence on the *creation* of data (Cook & Campbell, 1979). This is a *situational manipulation*. The transformation of variables into constants to exclude unwanted sources of variation and isolate operative variables can be deliberately carried out in laboratory conditions. However, both the statistical and the comparative methods have

no direct influence on the data. Control therefore occurs through *conceptual manipulation*. Investigators *select cases* either with similar values on a given property (which they want to keep constant) or different values on a given property.

MSSD and Comparable-Cases Control Strategies

In the comparative method matching as a control method plays a more important role than randomization. In the experimental and statistical methods randomization can be achieved more easily through manipulation and a large number of cases. When the number of cases is small randomization is more difficult. It means that cases are not enough to cover all the range of possible values on a given property or variable. The fact that matching is the main control technique in the comparative method has two main implications.

1. *Role of classification*. Because matching has such a central role in controlling unwanted sources of variation, in the comparative method conceptual treatment, classifications, and typologies become very important. The process of matching consists in *grouping cases according to similar values of given properties*. To keep constant a variable, all cases must have the same value on that variable. Thus, control through matching is gained by classifying and subclassifying (Smelser, 1976, pp. 168–169).
2. *"Most similar systems" design*. Consequently, for a number of authors comparative research designs are primarily designs in which cases are characterized by similarity. Lijphart (1975) argues "that it is more appropriate to reserve the term comparative method to the comparable-cases strategy and to assign the first solution [randomization] to the category of the statistical method" (p. 163). Cases are selected in such a way as to minimize the variance of control variables, and to maximize the variance of the experimental (independent and dependent) variables also to have a higher "degree of freedom."

Matching techniques were first developed in anthropology and introduced in sociology and political science as *methods of controlled comparisons* (Eggan, 1954; Hoenigswald, 1963), *specification* (Holt & Turner, 1970, p. 11) or *systematic comparative illustration* (Smelser, 1973, p. 53, 1976, p. 157). If researchers work with cases from a similar "area"

with a homogeneous culture and similar socioeconomic conditions, they can control more factors than if they would draw their cases from very different cultural and socioeconomic contexts. For this reason, these authors are in favor of *middle-range theories*, that is, research designs that are limited in their generalizability but allow controlled explanations.

The method of "controlled comparisons" was imported in comparative political studies by the famous book by Przeworski and Teune (1970), *The Logic of Comparative Social Inquiry* (see also Meckstroth, 1975). This type of research design takes a number of labels, the two most frequently used being "*most similar systems*" *design* (Przeworski & Teune, 1970) and *comparable-cases strategy* (George, 1979; Lijphart, 1975). In MSSD, researchers compare two or more cases that are as similar as possible to be able to focus on variation of the independent and dependent variables that constitute the relationship of interest.[17]

MSSD is a research *design*. It refers to the choice of cases and variables. With MSSD researchers proceed in a similar way as in controlled comparisons. They select cases from a homogeneous context, which allows one to minimize the number of "experimental variables" while increasing number of "control variables." The more circumstances the selected cases have in common, the stronger the leverage to identify which factor accounts for the variation of the dependent variable. The drawback is that risks of diffusion effects increase (see Galton's problem above).

The Role of Classification

The table on the next page clarifies the double role of classification for both *comparability* and *control* by taking as an example the (s)election of heads of state.

First, classification is indispensable for establishing what is *comparable*. If we are interested in the election of heads of state we must exclude cases in which these are not elected. Germany, Italy, and Switzerland *do not share* the attribute "election of the head of state" (value 0) with France, the United States, Austria, and so on (values 1). In Germany, Italy, and Switzerland, the head of state is appointed by the parliament whereas in France, the United States, Austria, and so on, the head of state is elected by the people. Therefore, Germany, Italy, and Switzerland are not comparable with France, the United States, and Austria on this specific property. If the level of generality is higher, however, and we use a more abstract concept on the ladder of generality (*se*lection rather than *e*lection), then cases become comparable. In all the cases there is a selection of heads of state (value 1).

The noncomparable cases would be countries in which heads of state are not selected (value 0), such as constitutional monarchies in which the office of head of state is hereditary (Britain, Sweden, Spain, Netherlands, etc.).

Role of Classification	*Level of Generality*			
	Lower Level		*Higher Level*	
Comparability (same attribute)	Election (1) No election (0)		Selection (1) No selection (0)	
Matching (same value)	Direct: Austria, France, Portugal, Ireland, Finland	Indirect: United States	Election: Austria, France, Portugal, Ireland, Finland, United States	Appointment: Germany, Italy, Switzerland

Second, classification is indispensable for *matching* variables. To transform a variable into a constant we take cases with the same value. In the case of the election of head of states, this variable has two values: direct and indirect election. In the case of the selection of head of states, this variable has two values: election or appointment. Note that at the lower level of generality "election" is the shared *attribute*, at the higher level it is a *value* of the nominal variable "selection." If we wish to see the impact of party fragmentation on governmental stability we may want to "control for" the type of selection (election or appointment) as the legitimacy of a directly elected head of state may compensate for the party fragmentation. Classification allows the creation of groups of homogeneous cases.